Adversarial Discriminative Domain Adaptation

Year: 2017 | Citations: 5000 | Authors: Eric Tzeng, Judy Hoffman, Kate Saenko, Trevor Darrell

---

Abstract

Adversarial learning methods are a promising approach to training robust deep networks, and can generate complex samples across diverse domains. They can also improve recognition despite the presence of domain shift or dataset bias: recent adversarial approaches to unsupervised domain adaptation reduce the difference between the training and test domain distributions and thus improve generalization performance. However, while generative adversarial networks (GANs) show compelling visualizations, they are not optimal on discriminative tasks and can be limited to smaller shifts. On the other hand, discriminative approaches can handle larger domain shifts, but impose tied weights on the model and do not exploit a GAN-based loss. In this work, we first outline a novel generalized framework for adversarial adaptation, which subsumes recent state-of-the-art approaches as special cases, and use this generalized view to better relate prior approaches. We then propose a previously unexplored instance of our general framework which combines discriminative modeling, untied weight sharing, and a GAN loss, which we call Adversarial Discriminative Domain Adaptation (ADDA). We show that ADDA is more effective yet considerably simpler than competing domain-adversarial methods, and demonstrate the promise of our approach by exceeding state-of-the-art unsupervised adaptation results on standard domain adaptation tasks as well as a difficult cross-modality object classification task.